Why Do We Need a Proper Geoid?

Petr VANÍČEK, Canada

Key words: gravity field, datum for heights, geoid, orthometric heights.

SUMMARY

This contribution enumerates some of the reasons for any country to have an appropriately computed, physically meaningful, geoid. In the first place, a physically meaningful vertical datum should guarantee the continuity of heights across national borders, continuity across the coastal zone, continuity on the continent and continuity globally! Also, nature has a habit of obeying physical laws. If an engineering project is designed by means of properly defined heights, including a properly defined datum (geoid), we can be assured that there will not be any nasty surprises, water will run down the hill, a lake surface will have the same height everywhere, etc. Using artificially defined heights (geodetic or orthometric heights referred to an artificial datum), there is no way of predicting their impact on the technical side of the project. The practice of using improper heights may also affect national boundaries on land and at sea, as it may leave questionable heights open to challenge by neighbouring countries.

Aside from the geodetic reasons, there are also scientific reasons for studying the gravity field of the Earth, including the geoid. The gravity field contains the most direct information about the mass density distribution within the Earth - think of ore deposits, oil, gas, etc. Should the state gather gravity data, or at least store and manage gravity data and the geoid as part of its mandate? Gravity field data are needed and used by scores of other sciences beside geodesy, such as oceanography, geophysics (both theoretical as well as exploration geophysics), satellite dynamics, space science, physics. These sciences are likely to expect government help with the data; they may be even looking to their respective governments for leadership. What kind of help will they get?

Why Do We Need a Proper Geoid?

Petr VANÍČEK, Canada

1. DATUM FOR HEIGHTS

Let me begin by stating my basic assumption that stands behind everything in this contribution. This assumption is that "We all agree that heights are a necessary component of surveying and mapping and that they should be as useful to the user as possible". To show what I mean by being "useful", let me cite from one of my earlier papers [Vaníček, 1998b]: "Should *ellipsoidal heights* - I prefer the term *geodetic heights*, since the term ellipsoidal heights, introduced to geodesy by our space brethren, implies the logical non sequitur (by analogy with the standard terminology of geoidal heights) "heights of the ellipsoid above the ellipsoid" – be used in everyday practice? Should the fact that a piece of a sea-shoreline with heights varying from -48.38 to -51.12 meters be perfectly all right with an engineer trying to come up with a plan for new port facilities in his area of interest? Should the idea be acceptable that an aqueduct can be adequately designed using heights which have no bearing on physical reality and which, consequently, cannot tell the designer which way is *up* and which way is *down*?"

The idea of useful heights is closely related to the physical meaning of heights. It is also closely related to the idea of a reasonable height datum. These two aspects cannot be always separated from one another. This contribution enumerates some of the reasons for any country to have an appropriately computed, physically meaningful, geoid, as their height datum. The alternative that is gaining some credence among surveyors as of late [Featherstone, 1995, 1998] is to replace a physically meaningful datum by a superficial surface obtained by fitting an arbitrary surface to a set of orthometric heights. Why should we have a physically meaningful vertical datum, rather than some conventional surface? To answer this question let us go back to the basic concepts [Vaníček and Krakiwsky, 1986] so we can understand what the problem is all about.

What is a datum and how is it defined anyway? Datum is a coordinate surface in a well defined coordinate system; *vertical datum* is a coordinate surface to which heights are referred. There are three kinds of vertical datums used in geodesy:

1) The geoid
2) The quasigeoid
3) The reference ellipsoid

The Geoid is nearly the universal choice for the vertical datum [Eckman, 1995; Veermer, 1998]. It is the datum for *orthometric heights* and for *dynamic heights*. We should recall that gravity is an essential part of the definition of any physically meaningful height (dynamic, orthometric or normal) definition as we shall discuss it below**.** The basis for defining a meaningful height is the following formula:

$$W_A = \sum_{i=0}^{A} g_i \Delta L_i \ , \qquad (1)$$

where W_A is the value of gravity potential at point A, g_i are the values of observed gravity along the levelling path and ΔL_i are the incremental height differences collected through levelling from the geoid to the point A. Geoid can be intuitively understood as it is an equipotential surface of the earth gravity field, i.e., it is the quintessential horizontal surface. The problem is that there are infinitely many equipotential surfaces.

The Quasigeoid takes the place of the geoid in Molodenskij's theory. It is a datum for *normal heights* which are not used outside of Russia, its former colonies and some countries in Europe.

The Reference Ellipsoid is a datum for *geodetic heights***.** It is not convenient for practice. As an example, the coast in eastern North America changes from –36 m in

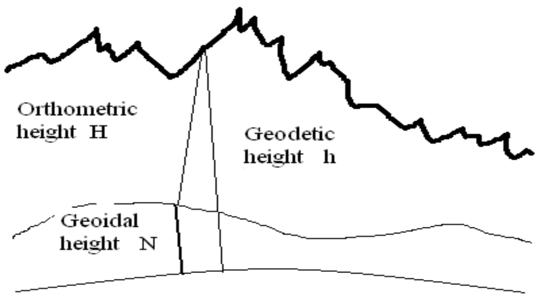

Figure 1 – Geodetic and orthometric heights

Chesapeake Bay to –13 m in Nova Scotia. But it is certainly needed for space-derived geodetic heights.

The geoid is defined as

$$W(\Omega) = W_0 = const., \qquad (2)$$

where Ω describes the position on Earth and the value of the constant may be chosen in either of two ways:

1) To equal a specific value of potential, $W_{0,}$
2) To make the geoid surface follow the *Mean Sea Level* (MSL) surface.

For the geoid as a height datum, the second definition is usually chosen. The reason is that it is easier to realise than the first definition and it provides a direct link to MSL. It is easier to access the MSL than it is to access a point of specific value of potential *W*.

What is MSL? Again two parallel definitions are used:

1) Point MSL – average of sea level at a point for a specific time epoch.
2) Surface MSL – conglomerate of all Point MSLs for a specific time epoch.

The separation between the geoid and MSL [for a certain time τ, i.e., MSL(τ)] is called *Sea Surface Topography*, and denoted as SST(τ). It is caused by sea dynamics, meteorological phenomena and other happenings. Its actual range is globally smaller than 4 metres. Note the parallel with the land topography.

In order to anchor the levelling network (network of points with heights associated with them) to a vertical datum, the vertical datum has to exist, i.e., it has to be realized in nature. This can be done in two different ways [Vaníček, 1990]:

1) Direct solution;
2) Indirect solution

The MSL(τ) as a function of time τ can be located at the shore. SST(τ) has to be derived by a special technique from different measurements, oceanographic, satellite, land levelling, etc. The MSL and SST together give us the height of a point on the geoid. This is shown on Fig 2. The SST has generally a negative sign in polar regions, positive sign in equatorial regions.

Figure 2. – Mean Sea Level and Sea Surface Topography

There are three available techniques for determining the SST:

1) Oceanic (steric) levelling;
2) Satellite altimetry combined with geoidal heights – needs geoidal height to determine geoidal height;
3) Zero frequency response analysis.

To discuss these techniques in detail would be beyond the scope of this paper. For further reading see, for instance [Sturges, 1967; 1974; Hamon and Greig, 1972; Merry and Vaníček, 1984]

A physically meaningful vertical datum guarantees the continuity of heights across borders, coastline zone, continuity in the continental context and global continuity [Vaníček, 1994; Jekeli, 2000]. Also, as nature has a habit of obeying physical laws, a physically meaningful vertical datum, combined with an appropriate definition of the height, we can be assured that in an engineering project there will be no nasty surprises: water will be running down the hill as it should, drilling at sea for oil and natural gas will take place on the own territory of the appropriate country as it should, a bridge built simultaneously from both shores of a river will meet in the middle, etc. Using artificial heights (geodetic heights or orthometric heights referred to a superficial datum), there is no way of predicting the consequences: anything can happen on account of the height selection.

2. PROPER HEIGHTS

What good would it do to use a proper datum for heights, yet have only approximate or inaccurate heights referred to it? It goes without saying that the datum should be complemented by proper and accurate heights. So let's have at least a brief look at the definition of proper heights. The best heights from the physical point of view are *dynamic heights* based on actual gravity. The dynamic height is defined as [Vaníček and Krakiwsky, 1986]

$$H_A^D = W_A / G \,, \qquad (3)$$

where G is a pre-selected value of gravity commonly used in the country. The somewhat foolishly perceived disadvantage of dynamic heights is that they do not describe heights as they would be measured by a rigid etalon. In fact, dynamic heights behave as if they were measured by a rubber meter, where the length of the meter changes in accordance with the strength of the gravity field. This behaviour is normally quoted as being responsible for the fact that dynamic heights are seldom used in practice even though they best describe the physics around us.

We have already mentioned the normal heights; the last but not the least member of the family of "physically meaningful" heights is the *orthometric height* defined [*ibid.*] as

$$H_A^O = W_A / \bar{g}_A' \,, \qquad (4)$$

where $\bar{g}_A'$ is the mean value of actual gravity between the point A and the geoid. Orthometric heights are not physically meaningful in the true sense of the term, but their geometrical meaning is well understood: an orthometric height is the distance between the geoid and the point of interest measured along the plumbline with a rigid etalon. Since the plumbline is slightly curved and twisted, the measurement cannot be visualised as being absolutely trivial, but the difference between the Euclidean distance and the length of the plubline is negligible for any practical purpose. Othometric heights do not adhere to physics in so far that the water may run up (an orthometric) hill, but they have an intuitive appeal to users and they are referred to the geoid as their datum. They also do not differ from dynamic heights very much. Consequently, they became the quintessential working heights and as such they are used in most parts of the world, either in their rigorous form or in their Helmert's approximation.

3. ACCURACY OF ORTHOMETRIC HEIGHTS

The next question is: Can orthometric heights be used in the adjustment of levelling networks? There were schools of thought that claimed that orthometric heights should not be used since they could not be adjusted, because they were ambiguous (their definition would allow to associate several values of orthometric height with any point). Such property is called *unholonomity* in physics. It has been shown by, e.g., Vaníček [1986] or Sanso and Vaníček [2006], that this is not the case: orthometric heights are holonomic and can indeed be properly adjusted.

Another objection to orthometric heights used to be that they could not be computed accurately because the density distribution within topography was not known; that all the practical formulae [based on Eq.(4)], used for computing orthometric heights were inaccurate. Recently, Tenzer et al. [2005] and Kingdon et al. [2005], showed that orthometric heights can be computed to a sufficient accuracy of, perhaps, a few millimetres, anywhere in the world.

Orthometric heights based on actual gravity seem to fit the requirements for accurate practical heights. On the other hand, the practice of using improper heights may give rise to all kinds of problems as discussed above; it may also affect national boundaries on land and at sea. Many boundaries are defined directly or indirectly, through sea or river shoreline (cf., concepts of high water, low water, etc.) and these are, of course, very much related to heights [Vaníček, 1994; 1998a]. For a proper definition of a boundary, it is essential that these shoreline points have correct heights associated with them to avoid any challenges from neighbouring countries and to insure repeatability. .

Properly defined orthometric heights are equivalent to bathymetric depths, even though they may be referred to different datums. Once the problem of different datums is eliminated, bathymetry and topography can be treated together. The unification of the concepts associated with heights and bathymetry would become a reality [Wells et al, 1996]. This unification presupposes that orthometric heights are properly defined and referred to the "real" geoid. If the orthometric heights were related to a superficial datum, this would not be true anymore.

4. SCIENTIFIC REASONS

Aside from the geodetic reasons discussed above, there are also scientific reasons for studying the gravity field of the Earth, including the geoid. The gravity field contains the most direct information about the mass density distribution within the Earth [Christou et al., 1989; Vajda and Vaníček, 1997]. Who needs to know anything about density distribution within the Earth?

Well, think of ore deposits, oil, gas, underground water, etc. Clearly, it is in anybody's interest to know as much about it as possible, as it is information about *national wealth*.

Should government get into the business of gathering gravity data, or at least storing and managing gravity data, including the geoid, as part of its mandate? In most of the developed countries, when it comes to various geo-data sets, their storage and management is considered an unquestionable responsibility of the governments of different levels; so why should gravity field data be treated differently?

Gravity field data are needed and used by scores of other sciences beside geodesy. As examples let's cite oceanography, geophysics (both theoretical as well as exploration), satellite dynamics, space science, physics [Vaníček and Christou, 1993, Nerem et al., 1995]. These are not, in general, just curiosity-driven sciences; the results coming from some of these scientific investigations have immediate practical applications. These sciences are likely to call on the various levels of government expecting to be helped with the data. They may be even looking to their respective governments for leadership. What kind of direction are they going to get?

REFERENCES

Christou, N., P. Vaníček and C. Ware, 1989. Geoid and density anomalies, *EOS*., 70 (22), pp. 625-631.

Ekman, M., 1995. What is the geoid? Finnish Geodetic Institute Report 95(4):49-51.

Featherstone, W.E., 1998. Do we need a gravimetric geoid or a model of the base of the Australian Height Datum to transform GPS heights? *The Australian Surveyor,* 43(4): 273-280.

Featherstone, W.E., 2002. Prospects for the Australian Height Datum and geoid model, in: Ádám J, Schwarz K-P (eds) *Vistas for Geodesy in the New Millennium*, Springer, Berlin Heidelberg New York, pp 96-101.

Hamon, B.V. and M.A. Greig, 1972. Mean sea level in relation to geodetic land levelling around Australia, *J Geophys Res* 77:7157-7162.

Jekeli, C., 2000. Heights, the geopotential, and vertical datums, Report 459, Dept of Civil and Environmental Engineering and Geodetic Science, Ohio State Univ, Columbus,

Kingdon, R., P. Vaníček, M. Santos, A. Ellmann and R. Tenzer, 2005. Toward an improved height system for Canada. *Geomatica*, Vol. 59, No. 3, pp. 241 to 249.

Merry, C.L. and P. Vaníček, 1983. Investigation of local variation of sea surface topography. *Marine Geodesy*, 7(1-4): 101-126.

Nerem, R.S., C.Jekeli and Wm. Kaula, 1995. Gravity field determination and characteristics: retrospective and prospective, *J Geophys Res* 100(B8):15053-15074.

Rapp, R.H., 1994. Separation between reference surfaces of selected vertical datums, Bull Géod 69(1):26-31.

Sanso, F. and P. Vaníček, 2006. The orthometric height and the holonomity problem. Presented at the Grafarend symposium in Stuttgart, February 18; *Journal of Geodesy* 80 (5), pp. 225-232.

Sturges, W., 1967. Slope of sea level along the Pacific Coast of the United States, *J Geophys Res* 72(14):3627-3637.
Sturges, W., 1974. Sea level slope along continental boundaries, *J Geophys Res* 79(6):825-830.
Tenzer, R , P. Vaníček, M. Santos , W. E. Featherstone, and M. Kuhn, 2005. Rigorous orthometric heights. *Journal of Geodesy* 79, pp. 1432-1394. DOI: 10.1007/s001-005-0445-2.
Vajda, P. and P. Vaníček, 1997. On gravity inversion for point mass anomalies by means of the truncated geoid. *Studia Geophysica et Geodaetica* , 41, pp. 329-344.
Vaníček, P., 1986. Letter to the Editor. *The Canadian Surveyor*, Vol. 40, No. 1, p. 53.
Vaníček, P., 1990. Vertical datum and the "NAD'88". Paper presented at ACSM/ASPRS annual convention, Denver, March 18-24. Also printed in *Surveying and Land Information Systems*, Vol. 51, No. 2, 1991, pp. 83-86.
Vaníček, P., 1994. On the global vertical datum and its role in maritime boundary demarcation; *Proceedings of international symposium INSMAP 94*, Hannover, Germany, September 19-23, pp. 243-250.
Vaníček, P., 1998a. On the errors in the delimitation of maritime spaces. *International Hydrographic Review* , LXXV(1), March, pp.59-64.
Vaníček, P., 1998b. The height of reason (a letter to the editor), *GPS World*, April 1998, p.14.
Vaníček, P. and E.J.Krakiwsky, second edition, 1986. *Geodesy: the concepts*, North Holland/Elsevier.
Vaníček ,P. and N.Christou (editors), 1993. *Geoid and its geophysical interpretations,* CRC Press, Boca Raton, Fla., USA. 343 pp.
Vermeer, M., 1998. The geoid as a product, Finnish Geodetic Institute Report 98(4):63-69.
Wells, D.E., A. Kleusberg and P. Vaníček, 1996. A seamless vertical-reference surface for acquisition, management and display (ECDIS) of hydrographic data, CHS Contract # IIHS4-122 Report for Canadian Hydrographic Survey, Ottawa, Technical Report # 179, GGE, UNB, pp. 73.

BIOGRAPHICAL NOTES

Petr Vaníček, P.Eng., Ph.D., Dr.Sc, is Professor Emeritus of geodesy in the Department of Geodesy and Geomatics Engineering at UNB. He retired in 1999, after 28 years of teaching and is now involved only in post-graduate student supervision and in research. His research interests cover the whole spectrum of geodesy, geophysics and applied mathematics. He is a fellow of AGU, IAG, Explorers Club, Senior Distinguished Scientist Humboldt awardee (1989), and recipient of CGU 1996 Tuzo J. Wilson medal. He is also author and co-author of about 450 publications including the comprehensive textbook "Geodesy: the concepts" used world-wide.
Personal web page: http://gge.unb.ca/Personnel/Vanicek/Vanicek.html.

CONTACT

Petr Vaníček
Department of Geodesy and Geomatics Engineering
University of New Brunswick
POB 4400, Fredericton, NB
CANADA E3B 5A3
Tel. + 1 506-453-5144,
Fax: + 1 506-453-4943
E-mail: vanicek@unb.ca
Web page: www.unb.ca/GGE/.